```
=> d his

        (FILE 'HOME' ENTERED AT 16:14:07 ON 26 JUN 2004)
        FILE 'CA' ENTERED AT 16:14:18 ON 26 JUN 2004
            E EFRON B/AU
L1      24 S E3-8
            E TIBSHIRANI R/AU
L2      30 S E3-7
L3       7 S L1-2 AND (BOOTSTRAP OR BOOT STRAP OR VALIDAT?)
L4  2552 S (CANCER? OR ADENOM? OR MALIGNAN? OR CARCINOM? OR
            TUMOR) AND (BOOTSTRAP OR BOOT STRAP OR VALIDAT?)
L5  1118 S L4 NOT PY>2000
L6   622 S (LEAVE ONE OUT OR LOO OR LOOCV) NOT LOOS
L7       6 S L5 AND (BOOTSTRAP OR BOOT STRAP)
L8       1 S L5 AND L6
L9      22 S L5 AND (CROSSVALIDA? OR CROSS VALIDAT? OR PREVALIDAT? OR PRE
            VALIDAT?)
L10      8 S L5 AND TRAIN?
L11      2 S L3 NOT PY>2000
L12     38 S L7-11
            FILE 'BIOSIS' ENTERED AT 16:34:17 ON 26 JUN 2004
L13    178 S L12
L14     53 S L13 AND (BOOTSTRAP OR BOOT STRAP)
            FILE 'MEDLINE' ENTERED AT 16:37:14 ON 26 JUN 2004
L15     46 S L14
            FILE 'CA, BIOSIS, MEDLINE' ENTERED AT 16:38:25 ON 26 JUN 2004
L16     92 DUP REM L12 L14 L15 (45 DUPLICATES REMOVED)


=> d bib,ab 1-92


L16   ANSWER 19 OF 92   BIOSIS   COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on
STN
AN    2000:123606   BIOSIS
TI    Bootstrap confidence intervals for the sensitivity of a quantitative
      diagnostic test.
AU    Platt, Robert W. [Reprint author]; Hanley, James A.; Yang, Hong
CS    Montreal Children's Hospital Research Institute, McGill University,
      2300 Tupper St, Montreal, QC, H3H 1P3, Canada
SO    Statistics in Medicine, (Feb. 15, 2000) Vol. 19, No. 3, pp. 313-322.
AB    We examine bootstrap approaches to the analysis of the sensitivity of
      quantitative diagnostic test data.  Methods exist for inference
      concerning the sensitivity of one or more tests for fixed levels of
      specificity, taking into account the variability in the sensitivity due
      to variability in the test values for normal subjects.  However,
      parametric methods do not adequately account for error, particularly
      when the data are non-normally distributed, and non-parametric methods
      have low power.  We implement bootstrap methods for confidence limits
      for the sensitivity of a test for a fixed specificity and demonstrate
      that under certain circumstances the bootstrap method gives more
      accurate confidence intervals than do other methods, while it performs
      at least as well as other methods in many standard situations.
```

AN   1999081515   MEDLINE
TI   Gamma-ray mutagen sensitivity and survival in patients with glioma.
AU   Sigurdson A J; Bondy M L; Hess K R; Toms S A; Kyritsis A P; Gu J; Wang
      L E; Wang X; Adatto P; Bruner J L; Yung W K; Levin V A; Wei Q
CS   Department of Epidemiology, The University of Texas M. D. Anderson
      Cancer Center, Houston 77030, USA.
SO   Clinical cancer research : an official journal of the American
      Association for Cancer Research, (1998 Dec) 4 (12) 3031-5.
AB   Despite advances in treatment of brain **tumors**, cerebral **malignant**
      gliomas are rapidly debilitating with poor survival. Patient age and
      tumor histology are known to influence survival in glioma patients, but
      these factors do not account for all of the variability in survival
      time. To identify additional useful predictors, we tested an assay
      that measures intrinsic gamma-ray mutagen sensitivity. Our hypothesis
      was that increased sensitivity of peripheral blood lymphocytes to
      chromatid breaks is associated with **tumor** aggressiveness and decreased
      patient survival. The eligible 76 patients were those with
      histologically confirmed **malignant** gliomas, seen at the University of
      Texas M. D. Anderson **Cancer** Center between 1994 and 1997, for whom we
      had sufficient blood for the in vitro gamma-radiation assay. After
      gamma-irradiation of each subject's lymphocytes, the frank chromatid
      breaks in 50 metaphases were averaged to calculate breaks/cell. On the
      basis of our patient series, we established a gamma-ray mutagen
      sensitivity cutoff point of 0.55 breaks/cell that was confirmed by
      bootstrap resampling techniques. Patients with values >0.55
      breaks/cell were considered sensitive. Kaplan-Meier and Cox
      proportional hazards modeling were used for the analysis. The gamma-
      ray mutagen-sensitive patients had worse survival than the nonsensitive
      patients, with an unadjusted hazard rate ratio of 1.6 (95% confidence
      interval, 0.9-2.8; P = 0.15). After adjustment for age, **tumor**
      histology, and extent of surgical resection, the hazard rate ratio was
      2.4 (95% confidence interval, 1.3-4.6; P = 0.0081). Our data suggest
      that gamma-ray mutagen sensitivity is a prognostic indicator of
      survival in glioma patients. The significance of these findings needs
      to be verified in studies with larger samples of patients with
      histologically

AN   1998:259416 BIOSIS
TI   Assessment of outcome prediction models for patients with localized
      prostate **carcinoma** managed with radical prostatectomy or external beam
      radiation therapy.
AU   D'Amico, Anthony [Reprint author]; Desjardin, April; Chung, Arnold;
      Chen, Ming-Hui; Schultz, Delray; Whittington, Richard; Malkowicz, S.
      Bruce; Wein, Alan; Tomaszewski, John E.; Renshaw, Andrew A.; Loughlin,
      Kevin; Richie, Jerome P.
CS   Joint Cent. Radiat. Therapy, Harvard Med. Sch., 330 Brookline Ave., 5th
      Floor, Boston, MA 02215, USA

AB   BACKGROUND. A clinical staging system for localized prostate **carcinoma**
      that provides reliable information on which management decisions
      regarding an individual patient can be based is lacking. This study
      compared the abilities of all published proposed clinical staging
      systems to predict time to prostate specific antigen (PSA) failure
      after radical prostatectomy or external beam radiation therapy for
      clinically localized prostate **carcinoma**. METHODS. A total of 1441
      clinically localized prostate **carcinoma** patients who were managed with
      radical prostatectomy at the University of Pennsylvania in Philadelphia
      (n = 688) or the Brigham and Women's Hospital in Boston (n = 288) or
      with external beam radiation therapy at the Joint Center for Radiation
      Therapy in Boston (n = 465) were entered into this study. Patients who
      received adjuvant or neoadjuvant hormonal or radiation therapy were
      excluded. Akaike's Information Criterion (AIC) and Schwartz Bayesian
      Criterion (SBC) estimates, which are comparative measures, were
      calculated for each clinical staging system. Pairwise comparisons of
      the AIC and SBC estimates for the most predictive clinical staging
      systems were performed using a formal **bootstrap** technique with 2000
      replications. RESULTS. Both the staging system based on the risk
      score and the staging system based on the calculated volume of prostate
      carcinoma and PSA (cVCa-PSA) optimized the prediction of time to
      posttreatment PSA failure. The cVCa-PSA system, however, provided a
      more clinically useful stratification of outcome. CONCLUSIONS.
      Improved clinical staging for patients with localized prostate
      carcinoma may be possible with parameters obtained during routine
      evaluation. **Validation** by other investigators is underway.

TI   COMPROC and CHECKNORM: Computer programs for comparing accuracies of
      diagnostic tests using ROC curves in the presence of verification bias.
AU   Zhou, Xiao-Hua [Reprint author]; Higgs, Richard E.
CS   Div. Biostatistics, Dep. Med., Indiana Univ. Sch. Med., Regenstrief
      Inst. Health Care, Riley Research Wing, RR 135, 702 Barnhill Drive,
      Indianapolis, IN 46202-5200, USA
AB   To assess relative accuracies of two diagnostic tests, we often compare
      the areas under the receiver operating characteristic (ROC) curves of
      these two tests in a paired design. Standard methods for analyzing
      data from a paired design require that every patient tested has the
      known disease status. In practice, however, some of the patients with
      test results may not have verified disease status. Any analysis using
      only verified cases may result in verification bias. COMPROC is an
      easy to use program for comparing the effectiveness of two diagnostic
      tests based on the area under the ROC curve in the presence of
      verification bias. COMPROC compensates for verification bias by
      implementing the maximum likelihood (ML) estimation of the areas and
      covariance matrix of two ROC curves under the missing at random (MAR)

assumption as described by Zhou (Biometrics 54 (1998) 349-366). This method assumes normality of the difference of the two ROC curve area estimators. We also describe a program CHECKNORM that does a **bootstrap** analysis to test this normality assumption (B. Efron, R.J. Tibshirani, An Introduction to the **Bootstrap**, Chapman and Hall, London, 1993). COMPROC allows for the inclusion of observed covariates that may influence the decision to verify the disease status of a patient. The program computes the estimates of the area under the ROC curve for the two diagnostic tests along with the variance of each area, the covariance between the two areas, a two-sided p-value, and a confidence interval for the difference of the areas. The programs COMPROC and CHECKNORM require the scripting language Perl and the statistical software SAS and can be run on both UNIX machines as well as PCs. The use of COMPROC and CHECKNORM is illustrated in a clinical study designed to compare relative accuracies of MRI and CT in detecting pancreatic **cancer**.

L16   ANSWER 48 OF 92   CA   COPYRIGHT 2004 ACS on STN
AN   130:107183   CA
TI   Estimating imprecision profiles in biochemical analysis
AU   Berweger, Christian D.; Muller-Plathe, Florian; Hanseler, Edgar; Keller, Herbert
CS   Laboratorium fur Physikalische Chemie, Eidgenossische Technische Hochschule Zurich, ETH Zentrum, Zurich, CH-8092, Switz.
SO   Clinica Chimica Acta (1998), 277(2), 107-125
AB   We describe a computer program IMPROFIL which dets. an imprecision profile of an anal. method from replicated measurements of samples. It calcs. the variance function, the coeff. of variation, the power of definition, the crit. limit, the limit of detection and the lower limit of quantification. The primary property, the variance function, is detd. by two alternative methods: the conventional max. approx. conditional likelihood method and the newly developed weighted abs. deviation method. For all quantities, confidence intervals are obtained using the **bootstrap** procedure. The program combines the use of robust numerical techniques, user-friendliness and integration into a spreadsheet program for data pre- and post-processing. The algorithms used are described in detail. Tests with synthetic data sets are used to **validate** the method and to establish its powers and limitations. Finally, its application to a practical anal. task (**tumor** marker CA 15-3 in human sera) is reported. For the method to yield a reliable est. of the variance function and the derived properties, certain min. requirements on the raw data must be met: They have to be spread throughout the concn. range of interest, there should not be less than three replicates per specimen, and there must be at least of the order of 25 (better at 50) specimens.

L16   ANSWER 50 OF 92   CA   COPYRIGHT 2004 ACS on STN
AN   129:271752   CA
TI   Identification of structural features and associated mechanisms of action for carcinogens in rats
AU   Cunningham, Albert R.; Klopman, Gilles; Rosenkranz, Herbert S.

CS   Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA, 15238, USA
AB   A set of chems. tested for carcinogenicity in rats that have been analyzed in the Carcinogenic Potency Database (CPDB) was subjected to CASE/MULTICASE (a computer-automated structure evaluation system) structure-activity relationship (SAR) analyses. This SAR system identifies structural features of chems. in a learning set that are assocd. with a predefined activity and produces an SAR model based on these characteristics. The rat CPDB used in this study consisted of 745 chems., 383 of which are carcinogens, 14 marginally active carcinogens (i.e., chems. that require a relatively high dose to induce carcinogenesis) and 348 are non-carcinogens. In an internal prediction anal. where CASE/MULTICASE 'predicted' the activity of chems. in the learning set, the system was able to achieve a concordance between exptl. and predicted results of 95%. This indicates that the program is able to adequately assess the chems. in the database. In a 10-fold **cross-validation** study where 10 disjoint sets of 10% of the chems. were removed from the database and the remaining 90% of the chems. were used as a learning set, CASE/MULTICASE was able to achieve a concordance between exptl. and predicted results of 64%. Using a modified **validation** process designed to investigate the predictivity of a more focused SAR model, the system was able to achieve a concordance of 71% between exptl. and predicted results. Among the major biophores identified by CASE/MULTICASE as assocd. with **cancer** causation in rats, several are derived from electrophilic or potentially electrophilic compds. (e.g., arom. amines, nitrogen mustards, isocyanates, epoxides). Other biophores however are derived from chems. seemingly devoid of actual or potential DNA-reactivity and as such may represent structural features of non-genotoxic carcinogens.

L16   ANSWER 53 OF 92 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN   1998:49030 BIOSIS
TI   Resampling and **cross-validation** techniques: A tool to reduce bias caused by model building.
AU   Schumacher, Martin [Reprint author]; Hollaender, Norbert; Sauerbrei, Willi
CS   Inst. Med. Biometry Informatics, Univ. Freiburg, Freiburg, Germany
AB   The process of model building involved in the analysis of many medical studies may lead to a considerable amount of over-optimism with respect to the predictive ability of the 'final' regression model. In this paper we illustrate this phenomenon in a simple cutpoint model and explore to what extent bias can be reduced by using **cross-validation** and **bootstrap** resampling. These computer intensive methods are compared to an ad hoc approach and to a heuristic method. Besides illustrating all proposals with the data from a breast **cancer** study we perform a simulation study in order to assess the quality of the methods.

AN   1997:261302   BIOSIS

TI   Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate **cancer**: A multi-institutional update.

AU   Partin, Alan W. [Reprint author]; Kattan, Michael W.; Subong, Eric N. P.; Walsh, Patrick C.; Wojno, Kirk J.; Oesterling, Joseph E.; Scardino, Peter T.; Pearson, J. D.

CS   James Buchanan Brady Urological Inst., Dep. Urol., Johns Hopkins Hosp., Baltimore, MD 21287-2101, USA

AB   Objective: To combine the clinical data from 3 academic institutions that serve as centers of excellence for the surgical treatment of clinically localized prostate **cancer** and develop a multi-institutional model combining serum prostate-specific antigen (PSA) level, clinical stage, and Gleason score to predict pathological stage for men with clinically localized prostate **cancer**. Design: In this update, we have combined clinical and pathological data for a group of 4133 men treated by several surgeons from 3 major academic urologic centers within the United States. Multinomial log-linear regression was performed for the simultaneous prediction of organ-confined disease, isolated capsular penetration, seminal vesicle involvement, or pelvic lymph node involvement. **Bootstrap** estimates of the predicted probabilities were used to develop nomograms to predict pathological stage. Additional **bootstrap** analyses were then obtained to **validate** the performance of the nomograms. Patients and Settings: A total of 4133 men who had undergone radical retropubic prostatectomy for clinically localized prostate **cancer** at The Johns Hopkins Hospital (n=3116), Baylor College of Medicine (n=782), and the University of Michigan School of Medicine (n=235) were enrolled into this study. None of the patients had received preoperative hormonal or radiation therapy. Outcome Measures: Simultaneous prediction of organ-confined disease, isolated capsular penetration, seminal vesicle involvement, or pelvic lymph node involvement using updated nomograms. Results: Prostate-specific antigen level, TNM clinical stage, and Gleason score contributed significantly to the prediction of pathological stage (P lt .001). **Bootstrap** estimates of the median and 95% confidence interval of the predicted probabilities are presented in the nomograms. For most cells in the nomograms, there is a greater than 25% probability of qualifying for more than one of the pathological stages. In the **validation** analyses, 72.4% of the time the nomograms correctly predicted the probability of a pathological stage to within 10% (organ-confined disease, 67.3%; isolated capsular penetration, 59.6%; seminal vesicle involvement, 79.6%; pelvic lymph node involvement, 82.9%). Conclusions: The data represent a multi-institutional modeling and **validation** of the clinical utility of combining PSA level measurement, clinical stage, and Gleason score to predict pathological stage for a group of men with localized prostate **cancer**. Clinicians can use these nomograms when counseling individual patents regarding the probability

of their **tumor** being a specific pathological stage; this will enable patients and physicians to make more informed treatment decisions based on the probability of a pathological stage, as well as risk tolerance and the values they place on various potential outcomes.

L16   ANSWER 59 OF 92   MEDLINE on STN
AN   97261687   MEDLINE
TI   The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis.
CM   Comment in: Med Decis Making. 1998 Jan-Mar;18(1):122-4. PubMed ID: 9456216
AU   Tourassi G D; Floyd C E
CS   Department of Radiology, Duke University Medical Center, Durham, North Carolina, USA.
SO   Medical decision making : an international journal of the Society for Medical Decision Making, (1997 Apr-Jun) 17 (2) 186-92.
AB   PURPOSE: To study the effect of data sampling on the predictive assessment of artificial neural networks (ANNs) for medical diagnostic tasks. METHODS: Three statistical techniques were used to evaluate the diagnostic performances of ANNs: 1) **cross validation**, 2) round robin, and 3) **bootstrap**. These techniques are different sampling plans designed to reduce the small-sample estimation bias and variance contributions. The study was based on two networks, one developed for the diagnosis of pulmonary embolism (1,064 cases) and the other developed for the diagnosis of breast **cancer** (206 cases). RESULTS: The three sampling techniques produced different performance estimates for both networks. The estimates varied substantially depending on the **training** sample size and the **training**-stopping criterion. CONCLUSION: The predictive assessment of ANNs in medical diagnosis can vary substantially based on the complexity of the problem, the data sampling technique, and the number of cases available.

L16   ANSWER 61 OF 92 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN   1996:486596 BIOSIS
TI   Computer-intensive statistical methods.
AU   **Efron, Bradley** [Reprint author]; **Tibshirani, Robert**
CS   Dep. Statistics, Sequoia Hall, Stanford Univ., Stanford, CA 94305, USA
SO   Armitage, P. [Editor]; David, H. A. [Editor]. (1996) pp. 131-147. Wiley Series in Probability and Statistics; Advances in biometry. Publisher: John Wiley and Sons, Inc., 605 Third Avenue, New York, New York 10158-0012, USA; John Wiley and Sons Ltd., Baffin Lane, Chichester PO 19 1UD, England.

L16   ANSWER 62 OF 92 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
AN   1997:18651 BIOSIS
DN   PREV199799317854
TI   **Bootstrap** confidence levels for phylogenetic trees. (Corrected and reprinted from Proceedings of the National Academy of Sciences, Volume 93, Issue 14, pages 7085-7090, 1996).

AU   **Efron, Bradley** [Reprint author]; Halloran, Elizabeth; Holmes, Susan

CS   Stanford Univ., Stanford, CA 94305, USA

AB   Evolutionary trees are often estimated from DNA or RNA sequence data. How much confidence should we have in the estimated trees? In 1985, Felsenstein (Felsenstein, J. (1985) Evolution 39, 783-791) suggested the use of the **bootstrap** to answer this question. Felsenstein's method, which in concept is a straightforward application of the **bootstrap**, is widely used, but has been criticized as biased in the genetics literature. This paper concerns the use of the **bootstrap** in the tree problem. We show that Felsenstein's method is not biased, but that it can be corrected to better agree with standard ideas of confidence levels and hypothesis testing. These corrections can be made by using the more elaborate **bootstrap** method presented here, at the expense of considerably more computation.

AN   1994:452660   BIOSIS

TI   **Bootstrap validation** of pharmacodynamic models defined via stepwise linear regression.

AU   Mick, Rosemarie [Reprint author]; Ratain, Mark J.

CS   Univ. Chicago Med. Cent., 5841 S. Mryland Ave. MC2115, Chicago, IL 60637, USA

AB   When a pharmacodynamic model is to be considered as the basis for individualized drug dosing, **validation** of the model is clearly warranted. Rigorous **validation** is problematic when the **training** data set to be modeled has too few data points and no independent test data set exists. A simulation method known as the **bootstrap** lends itself particularly well to this dilemma. **Bootstrap** sampling allows simulation of needed test data sets that mimic the initial data set. Model **validation** is then undertaken by repeating the model formulation procedure on the **bootstrap** samples. For illustration, a pharmacodynamic model for leukopenia was constructed by stepwise linear regression from data of 41 patients with **cancer** treated with the drug amonaride. Stepwise regression analyses were then repeated for 100 **bootstrap** samples, which verified the initial selection of covariates for the model. Next the regression parameters and residual error standard deviation of the model were repeatedly estimated for 200 additional **bootstrap** samples. The **bootstrap** results confirmed the initial formulation of the pharmacodynamic model from the **training** data set.

AN   1994:77745   BIOSIS

TI   The relative importance of prognostic factors in studies of survival.

AU   Schemper, M.

CS   Sec. Clin. Biometry, Dep. Med. Computer Sci., Vienna Univ.,
      Garnisongasse 13, A-1090 Vienna, Austria
AB   The relative importance of prognostic factors in regression can be
      measured either by standardized regression coefficients or by
      percentages of explained variation in a dependent variable. One
      advantage of using explained variation is the direct comparability of
      qualitative prognostic factors with others, or of groups of prognostic
      factors. The description of relative importance can be accomplished
      within marginal or partial effects analyses. It is demonstrated that
      it is possible not only to provide a descriptive ranking of prognostic
      factors according to their statistically determined importance, but
      also to make inferences concerning their relative importance, employing
      bootstrap techniques and procedures for multiple comparisons. The
      methods presented, which are new in the context of Cox regression, are
      exemplified by analyses of studies of lung **cancer** and breast **cancer**.

AN   1989:270124 BIOSIS
TI   A **BOOTSTRAP** ANALYSIS OF FOUR IN-VITRO SHORT-TERM TEST PERFORMANCES.
AU   BENIGNI R [Reprint author]
CS   LAB TOSSICOL COMP ECOTOSS, IST SUPERIORE SANITA, VIALE REGINA ELENA
      299, 00161 ROME, ITALY
AB   The present analysis is aimed at estimating the confidence intervals of
      a number of association measures that describe the relationships of 4
      in vitro short-term tests with rodent carcinogenicity, as well as with
      each other. The measures considered were: sensitivity, specificity and
      accuracy of the short-term tests with respect to chemical carcinogens,
      and performance dissimilarity indices (Hamming distances). The
      analysis refers to Salmonella, mouse lymphoma L5178Y cell mutation,
      chromosomal aberrations and sister-chromatid exchanges in Chinese
      hamster ovary cells, and is based on the data generated in the frame of
      the U.S. National Toxicology Program (NTP). It exploits the
      properties of a statistical technique, called **bootstrap**, to derive from
      only one sample of chemicals the variability intervals of the
      associations that the biological systems (mutagenicity assays and
      rodent carcinogenicity) would show in the 'universe' of the chemical
      compounds. The combination of the **bootstrap** technique with
      multivariate statistical methods pointed to a remarkable robustness and
      reliability of the information derived from the NTP data base, and
      provided descriptive insights into the data.

AN   1988:448137 BIOSIS
TI   **BOOTSTRAP** CONFIDENCE INTERVALS GOOD OR BAD?.
AU   **EFRON B** [Reprint author]
CS   DEP STATISTICS, SEQUOIA HALL, STANFORD UNIV, STANFORD, CALIF 94305, USA

```
=> log y
STN INTERNATIONAL LOGOFF AT 16:39:36 ON 26 JUN 2004
```